Versão 21-mai-05

O enigma do ENADE

Simon Schwartzman

Os resultados do Exame Nacional de Desempenho dos Estudantes, publicados pelo Ministério da Educação recentemente para 13 áreas de conhecimento, mostrou um resultado inesperado: em muitos cursos, a média de conhecimentos dos estudantes que concluíam os cursos em sua área específica de estudos era pior do que a média dos que entraram! O ENADE pretendeu inovar, em relação ao antigo Exame Nacional de Cursos, o Provão, em uma série de aspectos, entre os quais a realização de testes de conhecimentos gerais, além do teste de conhecimentos específicos; a aplicação da mesma prova para alunos no início e no fim dos cursos; e o uso de uma amostra de estudantes, no lugar da obrigação que os formandos tinham anteriormente de fazer o "provão". A última inovação foi o cálculo de um conceito geral de cada curso, a partir da ponderação destes diferentes resultados. Será que estas inovações resultaram em uma avaliação melhor?

Este texto não pretende fazer uma avaliação detalhada da metodologia e dos resultados do ENADE, mas, somente, chamar a atenção para alguns problemas e dúvidas mais evidentes, que precisariam ser melhor elucidados, a começar pela questão dos resultados negativos dos cursos. Esta nova versão se beneficia do fato de que já estão disponíveis, na página do INEP na Internet, vários documentos relativos ao ENADE, inclusive um resumo técnico geral, relatórios específicos por área de conhecimento, as provas utilizadas, e os gabaritos das respostas. Além disto, existe uma planilha Excel com os conceitos de cada curso, e mais algumas informações, que utilizei para tratar de entender melhor o que estava ocorrendo.

O resumo técnico explica os procedimentos de amostragem, assim como a estrutura geral das provas, compostas de uma parte comum, de avaliação da formação geral dos estudantes, composta de 10 itens, e outra sobre conhecimentos específicos de cada área, composta de mais 30 itens, de múltipla escolha e discursivos nos dois casos. Segundo o relatório, cada item das provas foi avaliado através da correlação que têm

com a prova como um todo, e foram eliminados os itens sem correlação. Existe muita informação também sobre o grau de dificuldade ou facilidade com de cada item.

A amostra

A parte estatística no resumo técnico está apresentada em uma linguagem de difícil compreensão para os não estatísticos, mas não é muito difícil de decifrar. Em total, o exame foi aplicado a uma amostra de 140 mil estudantes, para uma população de 250 mil, ou seja, 56% do total. O método foi de "amostragem estratificada com seleção aleatória simples em cada estrato", o que significa que os estudantes foram selecionados por sorteio dentro de todas as faculdades aonde existiam turmas iniciantes ou concluintes. O número de alunos selecionados dentro de cada turma dependeu de uma estimativa da variação das notas, feita a partir dos resultados do provão, quando havia. Quando as turmas eram muito pequenas, as provas eram aplicadas a todos os alunos.

Pode ser que este procedimento tenha reduzido o custo de analisar as perguntas abertas, mas, por outro lado, não deve ter havido ganho no trabalho de aplicação dos questionários, já que tanto faz aplicar uma prova para todos ou metade dos alunos de uma turma, e podem ter havido problemas de representatividade. Dez mil outros estudantes participaram do exame como voluntários, por mandatos de segurança, ou "como inscritos fora do prazo", mas não se sabe se os resultados das provas destes estudantes, que seguramente deformariam a amostra, estão incluídos ou não nos cálculos. Também não se sabe qual foi o rigor no sorteio dos alunos que participaram. Correm muitas histórias de cursos aonde só os melhores participaram, mas não é possível dizer se esta situação chegou a comprometer o resultado final das avaliações.

Como as provas foram elaboradas.

A prova de formação geral foi extremamente ambiciosa. Segundo o documento técnico, ela buscou avaliar "a formação de um profissional ético, competente e comprometido com a sociedade em que vive. Foram também consideradas as habilidades do estudante para analisar, sintetizar, criticar, deduzir, construir hipóteses, estabelecer relações, fazer comparações, detectar contradições, decidir, organizar, trabalhar em equipe e administrar conflitos." Isto em 8 questões de múltipla escolha. Além disto, haviam duas questões discursivas, que investigavam, segundo o resumo técnico, "aspectos como a clareza, a coerência, a coesão, as estratégias argumentativas, a utilização de vocabulário adequado e a correção gramatical do texto. Finalmente, na avaliação da formação geral foram contemplados temas como:

sociodiversidade, biodiversidade, globalização, novos mapas sociais, econômicos e geopolíticos, políticas públicas, redes sociais, relações interpessoais, inclusão e exclusão digital, cidadania e problemáticas contemporâneas."

Sem entrar no mérito dos conteúdos (que mistura questões comportamentais, como o "comprometimento com a sociedade", com habilidades gerais, como "clareza", e conhecimentos específicos, como "biodiversidade" e "novos mapas sociais"), é bastante improvável que esta prova tenha conseguido medir o que pretende, já que inclui cerca de 30 dimensões diferentes. Qualquer pessoa minimamente familiarizada com a elaboração de provas sabe que, para medir cada uma destas dimensões, seria necessário elaborar uma bateria de itens que precisariam ser testados de antemão, e resultariam em uma prova muito maior do que estes 10 itens que foram aplicados. Não é de se surpreender que o Ministério não esteja divulgando os resultados detalhados destas provas, em todas suas dimensões.

Que tipo de escala foi utilizado?

Os testes foram preparados por comissões de professores, e podem ser vistos na documentação do Ministério. A principal dúvida aqui é se estes testes obedecem a uma escala absoluta ou uma escala relativa. Na tradição brasileira, onde se usam escalas de notas de 0 a 10, a interpretação de sentido comum é que quem tira 10 é um excelente aluno, quem tira 7 é um aluno mais ou menos, quem tira 4 ou 5 está a ponto de ser reprovado, e quem tira menos do que isto não aprendeu o que devia. No entanto, basta pensar um pouco para entendermos que esta interpretação é totalmente subjetiva – alguns professores são bonzinhos, e dão notas altas para todos, e outros são "carrascos", e só dão notas baixas. Não é possível comparar um 7 do professor A com o professor B, ou um 7 em matemática com um 7 em geografia.

Existem duas soluções possíveis para este problema. A primeira, mais simples, consiste em transformar as notas em posições relativas em relação à média, de tal maneira que se possa dizer que tal aluno está bem acima da média, na média, abaixo da média, etc.; e depois a dispersão das notas é reduzida ou ampliada conforme a média dos desvios, ou o desvio padrão. É o que denomina "colocar as notas na curva", ou seja, dentro de uma distribuição que se aproxime de uma curva normal, que todos que sabem um pouco de estatística conhecem. É isto que fazia o Provão: cada curso era colocado em uma distribuição que sempre garantia que haveria uns poucos nos conceitos mais altos, outros tantos nos conceitos mais baixos, e os demais distribuídos ao longo da escala. Este procedimento permitia dizer quais eram os cursos melhores e piores, mas não permitia que comparássemos um "A" em nutrição

com um “A” em medicina, nem nos dizia qual era a nota mínima aceitável para cada curso.

A outra solução, muito mais difícil, consistiria em definir, com bastante clareza, qual o nível de conhecimentos esperado para os alunos de cada curso ao final de seus estudos. Esta metodologia é adotada, no Brasil, pelo SAEB, que é o sistema de avaliação da educação básica. O SAEB diz por exemplo que os alunos de matemática, na quarta série do ensino fundamental, devem ter uma pontuação de 250, os de 8ª série uma pontuação de 325, e os da terceira série do ensino médio, uma pontuação de 400, em uma escala de 0 a 500. A cada um destes valores está associada uma descrição bastante detalhada dos conhecimentos e competências que os estudantes devem ter em cada nível.

O ENADE tem um procedimento que se aproxima ao de uma distribuição relativa. Cada prova tem um conceito de 0 a 100, em separado □,para as questões de formação geral e de conhecimentos específicos. Depois, estas notas são recalculadas em função da média e do desvio padrão de todas as provas da área no país. Segundo o resumo técnico,

> O cálculo desse termo é realizado subtraindo-se da média das notas dos estudantes de uma instituição a média das notas de todos os estudantes do país, para cada uma das áreas, e dividindo-se o resultado da subtração pelo desvio-padrão das notas de todos os estudantes do país na área considerada. (p. 61).

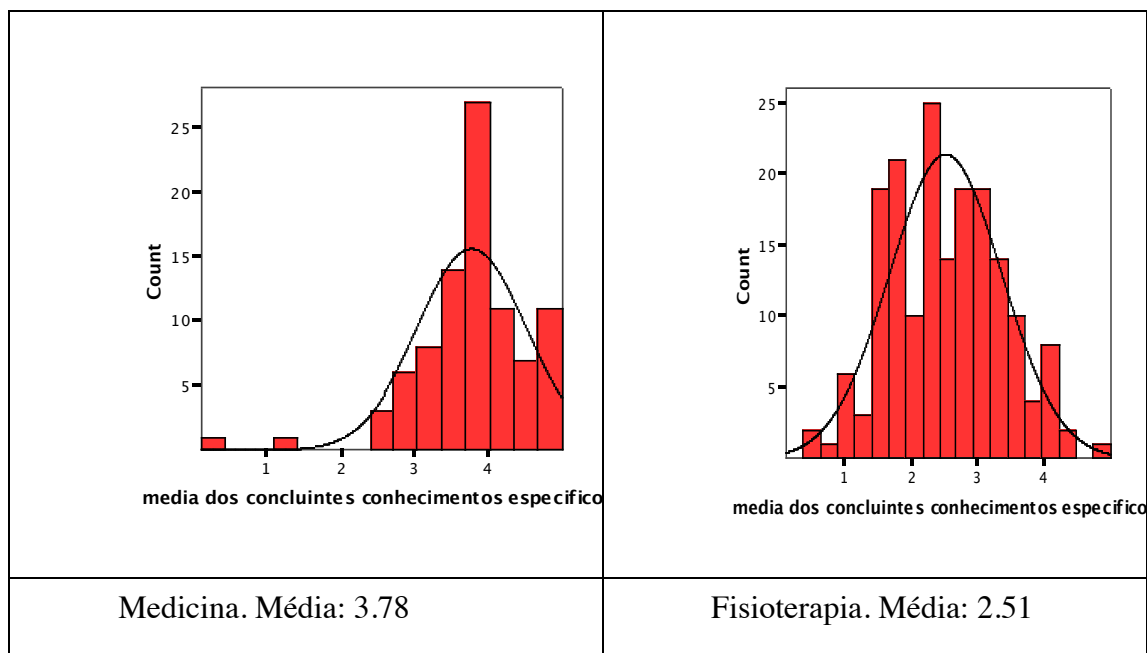
Assim, se a média de um curso for igual à média de todos os cursos, seu valor será zero. O resultado desta operação é fazer com que todos os cursos se distribuam ao redor da média zero, e com o desvio padrão de 1 (significando que, se a distribuição não for assimétrica, cerca de 67% dos cursos estarão um desvio padrão abaixo ou acima da média).

A partir daí, são feitas duas operações: primeiro, para que ninguém fique abaixo de zero na média, e possa variar entre zero e 5, diz o resumo técnico:

> Ao afastamento padronizado de cada uma das instituições será somado o valor absoluto do menor afastamento padronizado entre todas as instituições que oferecem o curso respectivo; em seguida, esse resultado será dividido pela soma do maior afastamento padronizado com o módulo do menor. Finalmente, o resultado desse quociente será multiplicado por 5 (p. 62).

O resultado desta operação é que a média, ao invés de ser sempre 2.5 para cada área de conhecimento, varia de área para área, sem que exista nenhuma interpretação clara

sobre o que significa esta variação[1]. As diferenças na distribuições das notas podem ser vistas, por exemplo, nos gráficos de medicina e fisioterapia, abaixo.

Medicina. Média: 3.78	Fisioterapia. Média: 2.51

Podemos concluir, por estes dados, que o cursos de medicina na média, são melhores do que os de fisioterapia? Claro que não. Para isto Ministério da Educação teria que ter estabelecido, com clareza, o nível de conhecimentos considerado ótimo, bom, razoável, inaceitável, etc., para cada área de conhecimento, e ter preparado os testes de maneira tal que houvesse uma correspondência clara entre as pontuações e estes níveis; e não poderia ter normalizado as notas em função das médias de cada área. Da forma em que está a interpretação dos resultados é semelhante à do Provão, no sentido de que o único que se pode dizer é a posição relativa de cada curso em relação aos demais da área, e sem uma distribuição normalizada dos resultados finais; com a agravante de que as posições são dadas em relação a uma média construída através de um procedimento arbitrário.

[1] "A forma de padronização utilizada na produção do conceito final depende crucialmente do tamanho do afastamento da menor nota em relação às demais. Assim sendo, esta padronização produz, em situações onde o menor conceito está muito isolado dos demais [exatamente o caso do curso de medicina, conforme o gráfico que está no seu texto], conceitos altos para a maioria dos outros. Isto explica a inflação de conceitos altos nos cursos de medicina e medicina veterinária." Agradeço a José Francisco Soares este comentário, que ajuda a entender melhor o problema.

O conceito final

Ministério tomou as três notas de que dispunha – a média de conhecimentos gerais, a de conhecimentos específicos iniciais, e a de conhecimentos específicos ao final do curso – e calculou uma média geral, usando um critério arbitrário de ponderação – 25%, 15% e 60% a cada um dos três resultados, respectivamente. Não se sabe de onde vieram estes pesos, e nem o que se supõe que esta combinação de critérios pretenda medir. Se o que interessa é informar à sociedade sobre os cursos que produzem pessoas mais bem formadas, o importante é o resultado da avaliação final, como fazia o Provão. Se o que interessa é o que o curso acrescenta, interessaria o ganho, se ele fosse uma medida confiável.

O resultado desta numerologia pode ser visto no quadro abaixo, com os resultados dos exames de conhecimentos específicos. As médias das áreas variam, tanto para os ingressantes, quanto para os concluintes, assim como para os dois grupos tomados em conjunto. Terapia ocupacional tem a média mais alta entre os concluintes, 4,08. mas isto não significa que, na média, os cursos desta área sejam melhores do que os de medicina, que têm média 3,78. A penúltima coluna mostra a diferença entre a pontuação final, depois que foram feitas as ponderações das três provas, e o conceito do exame final, que seria o adotado segundo o procedimento do Provão. O que se vê é que algumas áreas, como a terapia ocupacional e a farmácia, tiveram perdas de cerca de meio ponto, e outras foram menos afetadas. Supondo que a formação geral esteja correlacionada com a formação prévia dos alunos, que por sua vez está relacionada com a origem social das famílias dos alunos, é possível que esta ponderação esteja tendo o efeito de rebaixar os conceitos dos cursos que recebem alunos de origem social mais baixa.

Quadro 1

	Médias das provas de conhecimentos específicos				diferenças	
	ingressantes	concluintes	total	nota final	distorçao	ganhos
1 AGRONOMIA	2.94	2.87	2.91	2.74	-0.13	-0.07
2 EDUCACAO FISICA	2.77	2.92	2.83	2.92	0.00	0.15
3 ENFERMAGEM	3.09	2.90	3.02	3.00	0.10	-0.19
4 FARMACIA	2.23	3.31	2.69	2.90	-0.41	1.08
5 FISIOTERAPIA	3.45	2.51	3.09	2.45	-0.06	-0.94
6 FONOAUDIOLOGIA	2.68	2.52	2.60	2.40	-0.12	-0.16
7 MEDICINA	2.75	3.78	3.19	3.63	-0.15	1.03
8 MEDICINA VETERINARIA	3.15	3.55	3.32	3.42	-0.13	0.40
9 NUTRICAO	2.48	2.92	2.65	2.65	-0.27	0.44
10 ODONTOLOGIA	2.98	3.73	3.33	3.52	-0.21	0.75
11 SERVICO SOCIAL	2.62	2.71	2.66	2.57	-0.14	0.09
12 TERAPIA OCUPACIONAL	3.40	4.08	3.69	3.63	-0.45	0.68
13 ZOOTECNIA	3.33	2.29	2.88	2.57	0.28	-1.04
Total	2.90	3.04	2.95	2.93	-0.11	0.15

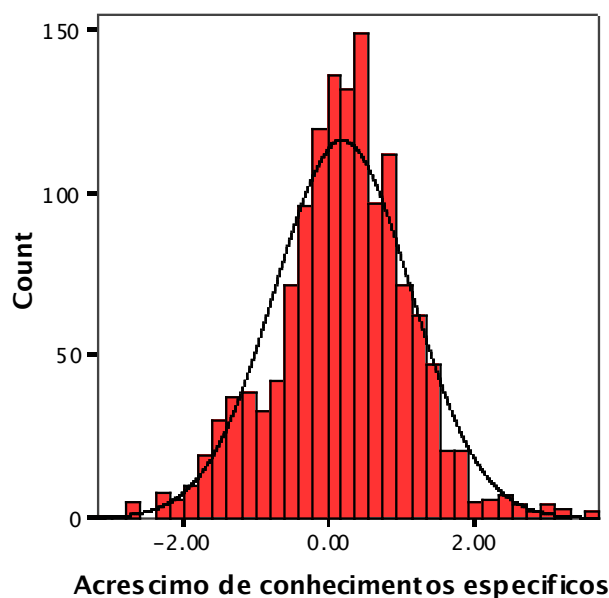
Depois, as médias finais foram agrupadas em cinco grupos, com conceitos de 1 a 5. Uma outra conseqüência da maneira em que as escalas foram construídas é que a metade dos cursos acabou ficando com conceitos 4 e 5, e somente 10% com 1 e 2. Isto não significa nada, mas é difícil transmitir para o público mais amplo, inclusive nas universidades, a interpretação correta destes dados. O mais provável é que as pessoas, simplesmente, multipliquem os conceitos por dois, e digam que metade dos cursos têm conceitos equivalentes aos tradicionais 8, 9 e 10 – uma conclusão totalmente falsa.

Quadro 2

Distribuição dos cursos das áreas por conceitos

	1	2	3	4	5	Total
1 AGRONOMIA	2.3%	8.1%	55.8%	29.1%	4.7%	100.0%
2 EDUCACAO FISICA	2.6%	1.8%	45.4%	44.5%	5.7%	100.0%
3 ENFERMAGEM	0.6%	5.0%	41.4%	43.6%	9.4%	100.0%
4 FARMACIA	0.7%	2.1%	58.9%	32.6%	5.7%	100.0%
5 FISIOTERAPIA	0.6%	26.6%	51.4%	17.9%	3.5%	100.0%
6 FONOAUDIOLOGIA	3.5%	22.8%	52.6%	19.3%	1.8%	100.0%
7 MEDICINA	1.1%		11.2%	58.4%	29.2%	100.0%
8 MEDICINA VETERINARIA	1.2%		16.5%	61.2%	21.2%	100.0%
9 NUTRICAO	1.9%	6.7%	64.8%	20.0%	6.7%	100.0%
10 ODONTOLOGIA	0.8%		9.4%	65.6%	24.2%	100.0%
11 SERVICO SOCIAL	14.0%	15.1%	21.5%	45.2%	4.3%	100.0%
12 TERAPIA OCUPACIONAL	7.4%		3.7%	48.1%	40.7%	100.0%
13 ZOOTECNIA		34.3%	34.3%	20.0%	11.4%	100.0%
	2.3%	8.1%	39.6%	39.5%	10.5%	100.0%

O antes e o depois

A aplicação da mesma prova para alunos iniciantes e concluintes dos cursos trouxe a grande surpresa do ENADE, o fato de que, em muitos cursos, os estudantes que concluíam sabiam menos do que os que entravam, nas questões específicas de cada área, como indicado na última coluna. O MEC só incluiu, na tabela Excel, a média dos conhecimentos gerais de todos os estudantes, sem distinguir se eram iniciantes ou concluintes, dizendo em seu relatório que não haviam diferenças importantes. Para as provas de conhecimento especifico, no entanto, foi possível calcular o "ganho", como a diferenças das médias dos iniciantes e concluintes. O gráfico abaixo mostra a distribuição dos resultados desta conta, para todos os cursos:

É uma bela curva normal, o que significa que os cursos se distribuem ao redor da média que é próxima de zero, com um número similar de cursos aonde os alunos saem melhor do que entram, e outros em que saem mais ignorantes, e com pouca diferença na maior parte deles. A interpretação mais óbvia para este resultado é que os exames, na realidade, não estão medindo o que se supõe que eles meçam. Pelo que diz o relatório técnico, os itens das provas foram submetidos a testes estatísticos para verificar sua capacidade de discriminação e sua correlação com o teste como um todo. Além desta validação interna, de consistência, testes e escalas precisam ser validados externamente, para saber se elas de fato estão medido aquilo que se supõe que elas medem. Não é apropriado, assim, usar uma balança para medir a altura das pessoas, mesmo que peso e altura costumem estar correlacionados. Neste caso, aparentemente, a única validação foi o consenso entre os professores que fizeram as provas. É possível imaginar que, em algumas áreas, os cursos são bem estruturados, e os professores que elaboraram as provas compartem uma visão mais definida do que se espera que os alunos aprendam ao longo do curso. Em outras áreas, no entanto, este consenso pode ser menos claro, e, na falta de um trabalho mais aprofundado de validação, elas podem estar medindo qualquer coisa, ou coisa nenhuma.

Quem ganha e quem perde?

Se a distribuição dos ganhos não é aleatória, que fatores poderiam explicar este resultado? Até o Ministério tornar disponível os microdados do ENADE, com informações individualizadas, e incluindo aí as respostas ao questionário

socioeconômico, a únicas informações que temos sobre os cursos são a área de conhecimento, o Estado ou região, e se o estabelecimento é federal, estadual ou privado. O quadro abaixo dá a média dos conceitos de conhecimento geral, conhecimento específico dos ingressantes, conhecimentos específicos dos concluintes, e ganhos – a diferença entre os conhecimentos dos que concluem e os que ingressam, por região, área de conhecimento e rede a que o curso pertence (as poucas municipais estão incluídas entre as privadas). A "significação", debaixo de cada grupo, diz se as diferenças dos conceitos se devem ao fato de os cursos pertencerem a determinados grupos, ou se são devidas a outras características dos cursos. Quanto menor o número, mais significativas são as diferenças entre grupos.[2]

Quadro 3

Médias dos resultados do ENAD				
	Geral	Inicial	Final	Ganho
Norte	2.35	2.63	2.63	0.05
Nordeste	2.59	2.82	3.02	0.28
Sudeste	2.56	2.93	3.04	0.16
Sul	2.74	3.00	3.21	0.22
Centro Oeste	2.51	2.71	2.74	0.12
Significação	*0.000*	*0.000*	*0.000*	*0.40*
Agronomia	2.27	2.94	2.87	-0.09
Educação Física	2.84	2.77	2.92	0.18
Enfermagem	3.00	3.09	2.90	-0.19
Farmácia	2.23	2.23	3.31	1.04
Fisioterapia	1.54	3.45	2.51	-0.94
Fonoaudiologia	1.76	2.68	2.52	-0.07
Medicina	3.64	2.75	3.78	1.08
Medicina Veterinária	3.18	3.15	3.55	0.29
Nutrição	2.06	2.48	2.92	0.41
Odontologia	3.26	2.98	3.73	0.73
Serviço Social	2.71	2.62	2.71	0.34
Terapia Ocupacional	3.23	3.40	4.08	0.76
Zoologia	2.77	3.33	2.29	-0.91
Significação	*0.000*	*0.000*	*0.000*	*0.00*
Estaduais	3.24	3.17	3.35	0.34
Federais	3.48	3.27	3.47	0.28
Privados	2.37	2.81	2.89	0.13
Significação	*0.000*	*0.000*	*0.000*	*0.01*

[2] Nota técnica: A medida usual utilizada nas análises de variância é o "f-ratio", que é tanto maior quanto mais importantes forem as diferenças entre grupos. Esta medida, no entanto, não é padronizada, e depende do número de grupos na análise, por isto optamos por indicar somente sua significação estatística. Uma significação de 0.000 indica que a probabilidade de que as diferenças entre grupos seja aleatória é menor do que 1 em mil, ou seja, as diferenças entre grupos são significativas; uma significação de 0.5 significa que existe uma probabilidade em 2 de que as diferenças entre grupos sejam aleatórias, ou seja, as diferenças entre grupos não são significativas.

Assim, a região em que o curso está afeta todos os resultados, menos o ganho dos alunos que terminam em relação os que iniciam. Era esperado que o Sudeste e Sul apresentassem valores mais altos nos três testes, e é interessante que o Nordeste não esteja tão longe dos demais.

Existem diferenças importantes nas médias entre as áreas de conhecimento: as médias mais altas são as de medicina, odontologia, terapia ocupacional e medicina veterinária, e as mais baixas, fonoaudiologia e fisioterapia. Mas, da forma pela qual foram feitas as padronizações, estes dados não são comparáveis, e não permitem dizer se uma área de conhecimento é melhor do que outra. O dado mais curioso, nesta parte, são os números negativos na última coluna. Há diferenças importantes de resultados: medicina, odontologia e terapia ocupacional acrescentam conhecimentos aos estudantes; zoologia e fisioterapia apresentam resultados muito piores na saída do que na entrada; e em muitas outras áreas, como a educação física, enfermagem, medicina veterinária, e fonoaudiologia, praticamente não há diferença. Para interpretar isto, no entanto é necessário primeiro decifrar o enigma do ENADE.

Finalmente, os efeitos das redes são os esperados, com os resultados melhores no sistema federal, e os piores no sistema privado, em média, com a ressalva que as universidades estaduais paulistas se recusaram a participar do exame.

Como utilizar os resultados obtidos?

Em principio, não teria sido necessário aplicar uma prova de conhecimentos gerais nos alunos iniciantes, já que esta informação, poderia ter sido obtida com os resultados do ENEM, o exame nacional do ensino médio, utilizado cada vez mais como informação para os exames vestibulares. O ENEM não pretende cobrir todos as 30 dimensões do exame preparado para o ENADE, mas, como a validade técnica deste exame do ENADE é muito duvidosa, utilizar o ENEM teria sido melhor, embora não fosse possível comparar os resultados na entrada e na saída dos cursos. Uma análise cuidadosa da formação geral dos alunos ao entrar nos cursos superiores poderia ser uma informação importante para interpretar os resultados obtidos ao longo dos cursos, junto com as demais informações obtidas pelo questionário socioeconômico. A comparação entre os resultados iniciais e finais neste teste poderia indicar se os cursos estão contribuindo para melhorar a formação dos alunos em algumas das 30 dimensões, mas é óbvio que isto jamais poderia ser feito com uma prova de 10 questões. O MEC reconheceu isto na prática, ao não divulgar os resultados de iniciantes e concluintes de forma separada, e ao não tentar analisar de forma separada as diferentes dimensões.

Quanto aos resultados específicos, um exemplo do que poderia ser feito é dado no quadro abaixo, que lista os 10 cursos de medicina que mais acrescentam conhecimentos a seus alunos, e os que menos acrescentam:

Quadro 4

Ganhos dos alunos dos cursos de medicina	Geral	Inicial	Final	Ganho
maiores ganhos:				
ESCOLA BAHIANA DE MEDICINA E SAUDE PUBLICA	2.7	0.1	3.8	3.7
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	4.3	2.2	5.0	2.8
ESCOLA DE CIENCIAS MEDICAS DE ALAGOAS	2.8	0.8	3.5	2.7
FACULDADE DE MEDICINA DO VALE DO ACO	3.0	2.5	5.0	2.5
UNIVERSIDADE FEDERAL DE GOIAS	4.4	2.6	4.9	2.3
UNIVERSIDADE ESTADUAL DE MARINGA	4.4	2.7	4.9	2.2
UNIVERSIDADE FEDERAL DE SAO PAULO	4.3	2.7	4.9	2.2
UNIVERSIDADE ESTADUAL DO OESTE DO PARANA	3.9	2.7	4.7	2.0
UNIVERSIDADE FEDERAL DE MATO GROSSO	4.4	2.5	4.4	1.9
ESCOLA DE MEDICINA SOUZA MARQUES DA FUNDACAO TECNICO-EDUCACI	3.5	2.3	4.1	1.8
Menores ganhos e perdas				
PONTIFICIA UNIVERSIDADE CATOLICA DE CAMPINAS	4.0	4.3	4.7	0.4
UNIVERSIDADE ESTADUAL DE MONTES CLAROS	5.0	3.5	3.8	0.3
UNIVERSIDADE DE MARILIA	3.0	2.3	2.5	0.2
UNIVERSIDADE FEDERAL DO ESPIRITO SANTO	4.6	3.8	4.0	0.2
UNIVERSIDADE LUTERANA DO BRASIL	3.3	2.8	3.0	0.2
UNIVERSIDADE DE RIBEIRAO PRETO	3.8	3.5	3.5	0.0
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	4.2	3.7	3.5	-0.2
UNIVERSIDADE FEDERAL DE PELOTAS	3.9	4.5	4.0	-0.5
UNIVERSIDADE DE CAXIAS DO SUL	4.1	3.7	3.1	-0.6
UNIVERSIDADE DE BRASILIA	3.5	3.1	1.3	-1.8
UNIVERSIDADE FEDERAL DE RORAIMA	2.8	5.0	2.6	-2.4

Estes dados podem ser sugestivos, mas não são confiáveis nem mesmo para a área de medicina, e devem ser menos confiáveis ainda para as demais. É improvável que os alunos que iniciam medicina em Roraima tenham aproveitamento máximo na avaliação, ou que os da Escola Baiana de Medicina e Saúde Pública tenham somente 0,1 de conceito.

A solução do mistério

Nos quadros 3 e 4, tentamos explorar o sentido dos ganhos e perdas de conhecimentos dos diferentes cursos, e chegamos a pensar que os resultados negativos poderiam ser explicados por problemas de validação das provas. Mas existe uma solução mais simples para o mistério, que só entendemos ao ler, na página 63 do resumo técnico, que *a padronização das notas do componente específico é feita separadamente para os ingressantes e concluintes, o que significa que os dois conceitos não são comparáveis*.

O correto teria sido padronizar todas as notas de medicina, por exemplo, em uma mesma escala normalizada, incluindo os iniciantes e os concluintes. Isto nos permitiria dizer por exemplo que, dentro de uma escala de 0 a 5 com média 2.5, os formados na faculdade X teriam conceito 4, e os ingressantes, conceito 1, o que daria

um ganho efetivo de 3. É assim que se procede no SAEB, por exemplo: existe uma escala única de matemática, e os resultados dos alunos das diversas séries são distribuídos ao longo desta escala.

Mas, se eu padronizo as notas dos iniciantes e concluintes *em separado*, não tenho como calcular este ganho, e o único que eu poderia fazer é comparar a posição relativa dos iniciantes e concluintes de cada curso em relação aos demais. Mesmo esta comparação seria difícil de interpretar, já que, como vimos, as médias variam arbitrariamente, em função das notas mais baixas que entram no ajuste feito para eliminar os valores negativos.

Se a padronização estivesse feita corretamente, as médias nacionais de desempenho para cada área deveriam ser as mesmas para ingressantes e para os concluintes. No entanto, a médias dos ingressantes e concluintes, quando tomados em sua totalidade, varia de área para área, mesmo quando ponderada pelo número de estudantes, como pode ser visto no quadro abaixo.

Quadro 5

	Ingressantes			concluintes		
	média ponderada	total	Std. Deviation	média ponderada	total	Std. Deviation
1 AGRONOMIA	3.1	5,910	1.05	2.98	3,616	0.74
2 EDUCACAO FISICA	2.7	35,761	0.64	2.93	17,534	0.69
3 ENFERMAGEM	3.0	35,553	0.43	2.79	11,207	0.82
4 FARMACIA	2.2	14,395	0.67	3.43	9,110	0.61
5 FISIOTERAPIA	3.4	23,302	0.49	2.42	11,872	0.75
6 FONOAUDIOLOGIA	2.6	2,271	0.54	2.53	2,089	0.86
7 MEDICINA	2.7	10,796	0.63	3.72	7,238	0.81
8 MEDICINA VETERINARIA	3.2	6,838	0.70	3.58	3,938	0.59
9 NUTRICAO	2.5	10,190	0.65	2.93	4,194	0.64
10 ODONTOLOGIA	3.0	8,034	0.81	3.72	6,881	0.51
11 SERVICO SOCIAL	2.5	8,660	1.02	2.56	4,845	1.09
12 TERAPIA OCUPACIONAL	3.4	1,057	0.78	4.07	712	0.86
13 ZOOTECNIA	3.4	1,510	0.68	2.34	894	1.17
Total	2.9	164,277	0.73	3.03	84,130	0.87

Esta é, pois, a solução do mistério: as diferenças entre iniciantes e concluintes não significam nada, porque as notas não comparáveis. Mas, se elas não são comparáveis, fica um mistério mais profundo ainda: porque foram feitas as duas provas?

Conclusão

Tudo indica que o ENADE foi feito de forma amadorística, com grande ambição, mas por pessoas com pouco ou nenhum conhecimento das metodologias adequadas para este tipo de trabalho. Avaliações em grande escala são procedimentos difíceis e complexos, e é normal que ainda exista muita coisa a ser feita para aperfeiçoar o ENADE. Os resultados teriam sido melhores se os ensinamentos adquiridos nas diversas rodadas do Provão tivessem sido aproveitados, e se o Ministério tivesse

buscado pessoas com conhecimento técnico e estatístico apropriados, na área de avaliação, para preparar e validar as provas.

O Ministério da Educação tem reiterado que os resultados do ENADE não têm muita importância neles mesmos, e precisariam ser vistos em conjunto com outras avaliações sobre as condições do curso, que estão prometidas para o futuro. Em parte, esta afirmação está relacionada com a idéia de que não se deve avaliar um curso somente pelos seus resultados, mas também pelo processo que levou a este resultado.

Isto depende, no entanto, do que se pretende com a avaliação. Fazendo uma metáfora com a medicina, é possível dizer que, para o paciente, o que interessa é se o hospital é bom e trata bem os doentes, e não se as instalações são ruins, ou se a verba é pouca, ou se os médicos não puderam ser bem formados. É claro que, para o médico que trabalha no hospital, é importante poder entender porque o tratamento não foi tão bom quanto poderia ser, e as autoridades de saúde devem ter o mesmo interesse, para melhorar a infra-estrutura e os processos que produzem os resultados. O Provão era um exame de resultados, para informar o público, enquanto que o ENADE é uma mistura, e a sua combinação com informações sobre processos e recursos, quando houver, vai obscurecer ainda mais a informação que o público e a sociedade necessitam, inclusive para poder apoiar as demandas por mais recursos e mais apoio pelos quais as instituições de ensino clamam.

No campo minado que é o debate sobre a educação no Brasil, muitas pessoas se opõem, por princípio, a medidas quantitativas em geral, e ao ordenamento de cursos em hierarquias de qualidade e desempenho em particular. Não foi fácil para o governo federal vencer estas resistências e dar continuidade à idéia de um Exame de Cursos, herdada do governo anterior, com a promessa de importantes aperfeiçoamentos. O risco é que as dificuldades de interpretação dos resultados do ENADE, ao invés de servir para aperfeiçoá-lo, acabem dando argumentos aos que nunca quiseram que ele existisse.